Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 6148–6153

On infinite dimensional linear programming approach to stochastic control ⋆

Maryam Kamgarpour * Tyler Summers **

* Automatic Control Laboratory, ETH Zürich, Switzerland (maryamk@ethz.ch)
** Mechanical Engineering, University of Texas at Dallas (tyler.summers@utdallas.edu)

Abstract: We consider the infinite dimensional linear programming (inf-LP) approach for solving stochastic control problems. The inf-LP corresponding to problems with uncountable state and input spaces is in general computationally intractable. By focusing on linear systems with quadratic cost (LQG), we establish a connection between this approach and the well-known Riccati LMIs. In particular, we show that the semidefinite programs known for the LQG problem can be derived from the pair of primal and dual inf-LPs. Furthermore, we establish a connection between multi-objective and chance constraint criteria and the inf-LP formulation.

Keywords: stochastic control, linear programming, semidefinite programming

1. INTRODUCTION

Optimal control of discrete time stochastic systems can be addressed via the dynamic programming (DP) (Bellman, 1957) principle of optimality. For an infinite horizon average or discounted cost problem, the optimal cost function and control policy can be computed as the fixed point of the so-called dynamic programming operator. In general, computing this fixed point is challenging and thus, several approximate approaches based on the DP principle of optimality have been developed.

An alternative approach to solving stochastic control problems is linear programming (LP) (Puterman, 2009; Hernández-Lerma and Lasserre, 1996). If the control and input spaces are uncountable, the corresponding LP is infinite dimensional (inf-LP). In the primal form of this LP, the optimization variable is the *occupation measure*, which measures infinite horizon occupancy of state and inputs in each Borel subset of the product state input space. An optimal policy may be derived from the optimal occupation measure, while the optimal value function is the optimizer of the dual of this LP.

In addition to providing an elegant alternative formulation of the optimality conditions for a stochastic control solution, in the LP approach constraints have a natural interpretation. By properly constraining the occupation measure, one can ensure probabilistic constraints on the state trajectory or can ensure bounds on multiple objectives. Such formulations of constrained stochastic control were considered in (Borkar, 1994; Feinberg and Shwartz, 1996; Altman, 1999; Hernández-Lerma and González-Hernández, 2000; Hernández-Lerma et al., 2003).

The inf-LP formulation is in general computationally intractable. For problems with polynomial data, this inf-LP can be approximated via a sequence of semidefinite programs (SDPs) (Savorgnan et al., 2009; Summers et al., 2013). These recent works are among the few that explore the inf-LP approach for computation of optimal value function and policies in a stochastic control problem.

The abstract inf-LP work has not attempted to establish clear connections with the well known, computationally tractable Linear Matrix Inequality (LMI) formulations of optimal control. In particular, for a stochastic linear system with quadratic cost (LQG), one can formulate the so-called Riccati LMI to find the optimal value function of the LQG problem (Boyd et al., 1994; Balakrishnan and Vandenberghe, 2003). Similarly, the well known LMI formulations have not attempted to show how these results can be derived from a more general approach to stochastic optimal control, namely the inf-LP approach.

In this work, we establish the connection between the inf-LP approach and the well-known Riccati LMIs for LQG problems. This inf-LP in general, includes infinitely many constraints on the occupation measure. The relaxation of these constraints to moments up to order two of the occupation measure and taking the dual of this problem results in the well-known Riccati LMI solution approaches. Since the variables in the relaxation of primal inf-LP are discounted moments of the state and input, moment constraints or certain class of chance constraints can be naturally encoded in the inf-LP formulation.

Our paper is organized as follows. In Section 2 we review the inf-LP approach to discrete-time infinite horizon discounted stochastic control. In Section 3 we apply the approach to LQG problems. In Section 4 we provide numerical case studies. In Section 5 we summarize the results.

⋆ This research is partially supported by M. Kamgarpour's European Union ERC Starting Grant, CONENE and by T. Summers' the US National Science Foundation under grant CNS-1566127.

10.1016/j.ifacol.2017.08.979

2. INF-LP APPROACH TO STOCHASTIC CONTROL

Consider the discrete-time stochastic system

$$x_{t+1} \sim \tau(B_x|x_t, u_t), \tag{1}$$

where $x_t \in X$, $u_t \in U$, and $\tau(.|x,u)$ is a stochastic kernel. It assigns a probability distribution to $B_x \in \mathcal{B}(X)$ given x and u, where $\mathcal{B}(X)$ is the set of Borel subsets of X. The stochastic control problem is defined by

$$\min_{\pi \in \Pi} \mathbb{E}^{\pi}_{\nu_0} \sum_{t=0}^{\infty} \alpha^t c_0(x_t, u_t). \tag{2}$$

Above, $c_0 : X \times U \to \mathbb{R}_+$ is the running cost and $\alpha \in (0,1)$ is a discount factor, ν_0 is an initial state distribution. We consider randomized policies $\pi \in \Pi$, where Π is the set of probability measures on U given X. That is, for each $x \in X$, $\pi(x)$ gives a probability distribution on the input space U. The expectation $\mathbb{E}$ is with respect to the probability measure induced by ν_0, π and τ.

The solution to the stochastic control problem above can be characterized as the solution of an infinite dimensional linear program (inf-LP). To present this inf-LP, we first define the infinite dimensional optimization spaces for the primal and dual LPs. Define the weight functions

$$w(x,u) = \epsilon + c_0(x,u), \;\; \tilde{w}(x) = \min_{u \in U} w(x,u), \tag{3}$$

where $\epsilon > 0$ so that the weights are bounded away from zero. Let $\mathcal{F}(X \times U), \mathcal{F}(X)$ denote the space of real valued measurable functions with bounded $w, \tilde{w}$-norms, respectively. That is, for $f \in \mathcal{F}(X \times U), \tilde{f} \in \mathcal{F}(X)$:

$$\sup_{(x,u)} \frac{|f(x,u)|}{w(x,u)} < \infty, \quad \sup_{x} \frac{|\tilde{f}(x)|}{\tilde{w}(x)} < \infty.$$

Let $\mathcal{M}(X \times U), \mathcal{M}(X)$ denote the space of measures with finite $w, \tilde{w}$-variations, respectively. That is, for $\mu \in \mathcal{M}(X \times U), \tilde{\mu} \in \mathcal{M}(X)$:

$$\int_{X \times U} w d\mu < \infty, \quad \int_X \tilde{w} d\tilde{\mu} < \infty. \tag{4}$$

Define the linear map $T : \mathcal{M}(X \times U) \to \mathcal{M}(X)$ as:

$$[T\mu](B) = \tilde{\mu}(B) - \alpha \int_{X \times U} \tau(B|x,u)\mu(dx,du), \tag{5}$$

where $\tilde{\mu}(B) := \mu(B,U)$ and $B \in \mathcal{B}(X)$. Analogously, define the linear map $T^* : \mathcal{F}(X) \to \mathcal{F}(X \times U)$ as:

$$[T^* v]\,(x,u) = v(x) - \alpha \int_X \tau(dy|x,u)v(y).$$

Note that the second term above $\int_X \tau(dy|x,u)v(y)$, is the expectation of the function v under the stochastic kernel τ. One can verify that T and T^* are adjoint operators:

$$< T^* v, \mu >_{X \times U} = < v, T\mu >_X,$$

where the bilinear maps are given by:

$$< c, \mu >_{X \times U} = \int_{X \times U} c(x,u)\mu(dx,du),$$
$$< v, \nu >_X = \int_X v(x)\nu(dx).$$

In the remainder, for simplicity, we drop the subscript of $< . \, , \, . >$ since the space is clear from the context. To formulate the inf-LP corresponding to stochastic control, we need the following standard assumptions (Hernández-Lerma and Lasserre, 1996).

Assumption 1.

(a) The cost c_0 is lower semi-continuous and inf-compact, that is, for every $x \in X$, $r \in \mathbb{R}$, the set $\{u \in U \mid c_0(x,u) \leq r\}$ is non-empty and compact.
(b) The stochastic kernel τ is weakly continuous.
(c) $\sup_{X \times U} \int_X \tilde{w}(y)\tau(dy|x,u)/w(x,u) < \infty$.
(d) $\nu_0 \in \mathcal{M}_+(X)$.

Let $\mathcal{M}_+(X \times U) \subset \mathcal{M}(X \times U)$ denote the cone of non-negative measures. For $\nu_0 \in \mathcal{M}_+(X)$, the constraint on $\mu \in \mathcal{M}(X \times U)$, denoted by $\nu_0 - T\mu = 0$ refers to

$$\nu_0(B_x) - [T\mu](B_x) = 0, \quad \forall B_x \in \mathcal{B}(X). \tag{6}$$

Theorem 1. The stochastic control problem (1), (2) can be equivalently formulated as the following inf-LP:

$$\min_{\mu \in \mathcal{M}_+(K)} \quad < c_0, \mu > \tag{P-SC}$$
$$\text{s.t.} \quad \nu_0 - T\mu = 0. \tag{7}$$

We summarize the idea of the proof and refer the readers to (Hernández-Lerma and Lasserre, 1996) for details. Given a policy $\pi \in \Pi$, one can define $\mu \in \mathcal{M}_+(X \times U)$ as

$$\mu(B_x, B_u) = \sum_{t=0}^{\infty} \alpha^t \mathbb{P}^{\pi}_{\nu_0}\{(x_t, u_t) \in (B_x, B_u)\}, \tag{8}$$

where $B_x \in \mathcal{B}(X), B_u \in \mathcal{B}(U)$. This measure corresponds to discounted probability of (x_t, u_t) being in any Borel subset of $X \times U$ and is referred to as the occupation measure. It can be verified that the occupation measure satisfies $\nu_0 - T\mu = 0$. Furthermore, given any $\mu \in \mathcal{M}_+(X \times U)$, there exists a policy $\varphi \in \Pi$, satisfying

$$\mu(B_x, B_u) = \int_{B_x} \varphi(B_u|x)\tilde{\mu}(dx), \tag{9}$$

for all $B_x \in \mathcal{B}(X), B_u \in \mathcal{B}(U)$ [Proposition D.8(a) in (Hernández-Lerma and Lasserre, 1996)]. It can be shown that the cost (2) corresponding to the policy φ is

$$\mathbb{E}^{\varphi}_{\nu_0} \sum_{t=0}^{\infty} \alpha^t c_0(x_t, u_t) = < c_0, \mu > . \tag{10}$$

Putting the above results together, the problem of finding the optimal policy for (2) can be equivalently formulated as finding a measure minimizing (10) subject to (7).

Whereas the inf-LP above provides the optimal occupation measure and the optimal policy for the stochastic control problem, the dual of this inf-LP can be used to find the optimal value function. Furthermore, the duality gap is zero (Hernández-Lerma and Lasserre, 1996).

To define this dual inf-LP, let the constraint on $v \in \mathcal{F}(X)$, denoted by $c_0 - T^* v \geq 0$ refer to

$$c_0(x,u) - [T^* v](x,u) \geq 0, \; \forall (x,u) \in X \times U. \tag{11}$$

The dual inf-LP is given by:

$$\max_{v \in \mathcal{F}(X)} \quad < v, \nu_0 > \tag{D-SC}$$
$$\text{s.t.} \quad c_0 - T^* v \geq 0. \tag{12}$$

Remark. Constraint (12) is the Bellman inequality. In particular, based on the Bellman principle of optimality, a function v^* is the optimal value function of the stochastic control if and only if $c_0 - T^* v = 0$. Thus, the optimizer of the above inf-LP satisfies the Bellman equality.